EOSDIS IV&V Technical Analysis Memorandum (TAM)

To: (Individual)

From: EOSDIS IV&V Team

Subject: Interactive Cost Model

1. *Context*

This memo is to discuss the Interactive Cost Model provided to NASA by Ken Bowyer of HAIS. The model is a Microsoft Excel based spreadsheet that allows the user to vary the input of 15 separate variables. The first three variables the user must input are (1) **Launch_Date**, (2) **L0_Data**, (level zero data) and (3) **Site**, see table 1. The other 12 variables consist of the Processing and Data Volume Requirements for the years 1997 through 2002. The user can vary the amount of Processing Load Effective MFLOPS and Data Volume GB/Day over this time period.

The input areas are at the top of the spreadsheet and are defined by bold outline around each input box. The model calculates the effect of changing these variables, allowing the user to construct "What if" scenarios for different processing requirements. The output from this model is "Cost by Fiscal Year" table showing the costs from 1994 through 2003 for Processing, RAID Disk, Archive, Maintenance, and a Total Cost by fiscal year, see table 2.

Table 1. Data Entry Fields

Miscellaneous Requirements		
Launch_Date	8/16/93	m/d/yy
L0_Data	0.090	GB/Day
Site	LARC	DAAC Name

Processing and Data Volume Requirements		
Calendar Year	Processing Load Effective MFLOPS	Data Volume GB/Day
1997	13.894	5.275
1998	905.911	11.339
1999	905.911	11.339
2000	905.911	11.339
2001		
2002		
Peak Values	905.911	11.339

Context cont.

Table 2. Cost by Calendar Year

Cost by Fiscal Year					
Fiscal Year	**Processing**	**RAID Disk**	**Archive**	**Maintenance**	**Totals**
1994	$0	$0	$0	$0	$0
1995	$7,785	$5,541	$0	$0	$13,326
1996	$18,451	$11,968	$0	$843	$31,262
1997	$0	$0	$70,418	$3,006	$73,424
1998	$1,819,988	$24,036	$98,011	$7,859	$1,949,894
1999	$1,318,109	$21,147	$93,110	$120,343	$1,552,710
2000	$0	$0	$6,942	$232,582	$239,524
2001	$0	$0	$0	$257,952	$257,952
2002	$0	$0	$0	$258,051	$258,051
2003	$0	$0	$0	$64,513	$64,513
Totals	$3,164,334	$62,693	$268,481	$945,149	$4,440,657

Discussion

After a brief analysis of the model the following were identified as areas of concern:

1. The Model was submitted to Dr. H. K. Ramapriyan on March 7, 1995. However, the model was constructed in such a way that it did not function properly on our Intel-based PC's which were using a higher version of MS Excel. After other people at GSFC had the same difficulty, an updated model was provided.

2. In an introductory paragraph to Dr. Ramapriyan, the creator of the model, Ken Bowyer states "The current required input values in the model are based on the TRMM CERES instrument requirements from the AHWGP data." Consequently, the input values for Processing Load Effective MFLOPS and Data Volume GB/Day are zero for the years 2001 and 2002. It would appear that NASA would want an interactive cost model to reflect the requirements for the EOSDIS, not TRMM CERES. The model does make provisions for handling data in the years 2001 and 2002.

3. Provided with the model are four pages of text which include the Purpose, Assumptions, Inputs, Outputs for the model. In the Assumptions section there are detailed explanations of the methods used to derive key figures and there impact on the Price/Performance curves that are used. Yet, there are no quantifiable numbers given which would allow duplication and verification of these assumptions.

For example, in the Assumptions section 1a. "The Processing Price/Performance curve has a slope in which the Price/MFLOP decreases by 21% per year. The starting point for this curve is based on the average Price/MFLOP of an SGI Power Challenge with 2 processors and a DEC 7000/620 with 2 processors."

3. **cont**. Both of these machines can be purchased in a number of different configurations. Each configuration can significantly effect that machine's processing speed and efficiency. Yet, the projected performance of each machine, (which would allow a more complete analysis of this proposal,) are not defined in numerical terms. The reader is left to calculate a performance level for these machines and project a 21% decrease in the Price/Performance curve. The reader is also left to assume that this 21% decrease is through the end of the contract, October 31, 2002, rather than December 31, 2000 where the initial input values end.

4. Under Assumptions 1b. it refers to the RAID Disk and the storage capacities of the units selected There are two sizes of RAID Disks mentioned and the Price/Performance curve that is selected based upon the required RAID Disk capacity. The writer then states, "The starting point for this curve is based on the average Price/GB of Maximum Strategy and IBM RAID Disk".

The "Maximum Strategy and IBM RAID Disk" is not identified in the text accompanying the model.

5. Under Assumptions 1d. it states, "All curves add a 36% tax to the price/performance to account for "Push" side COTS communications cost and contractual burdens". This would imply that 36% is added to the total cost, in addition to:

> Assumptions section 1a. "A 22% tax is added to the Price/MFLOP to account for processing COTS software cost."
>
> Assumptions section 1c. "A 20% tax added to the Price/TB to account for file storage management COTS software cost."

Thus, the reader is left to determine **if** and **when** each of the following "taxes" are added and if they are cumulative.

6. Under the Inputs section 1c it states: "The DAAC site that will archive the data is entered (ASF, EDC, GSFC, JPL, LARC, MSFC, or NSIDC are acceptable inputs)". It appears in the coding of this model that only three sites are valid, EDC, GSFC, LARC. Based on preliminary test data from the model, the costs remain identical as the site changes and the L0_Data is increased from .09 to 1,000. It is unclear why the user must input a specific site when it appears to have no impact on the model from a cost standpoint. Therefore, it appears that the costs are assumed to remain constant, regardless of the site selected. See attached graphs.

2. *Discussion cont.*

7. The following represent the starting point for the calculation of the various curves and cost identified within the model. If Intermetrics is asked to comment on the amounts stated in the categories in terms of there reasonableness, Intermetrics would need HAIS to identify the basis and rational for selecting these starting points.

a. The processing price/performance curves for the price of processing per MFLOP.

Processing Price/Performance Parameters		
YEAR	Units	Run-Values
1994	DOL_MFLOP_94	591.072126434503

b. The starting point for the small (30 GB) and large (100 GB) Raid Disk

YEAR	Units	Run-Values	Units	Run-Values
1994	DOL_GB_LRG_94	8689.2254889375	DOL_GB_SML_94	9562.855705

c. The starting point for Archive Price/Performance curves for small (10 TB) and large (100 TB) archive storage units. Please note two curves start in 1994 and run through 1999 and two curves start in 2000 and run through 2002.

YEAR	Units	Run-Values	Units	Run-Values
1994	DOL_TB_LRG_94	43611.6381033998	DOL_TB_SML_94	104500.073382682
2000	DOL_TB_LRG_00	2515.95277379804	DOL_TB_SML_00	6028.60293543435

3. *Recommendations*

The preceding was a preliminary analysis of the Interactive Cost Model. It does not address the underlying coding which drive the calculations or give a detailed analysis of the model's output to determine if it is possible or probable. It appears there are enough unknown, or unclear areas within the model, its Assumptions, Inputs and Outputs to warrant a more thorough analysis.

A complete explanation of the Assumptions section is needed to determine if the "taxes" upon "taxes" methods of cost build up are rational and reasonable. Also an explanation of the starting points that were used for the performance curves is necessary, see discussion 7.

The model is password protected which makes decomposing it difficult. The text boxes within the model are too small, and the explanations contained within them are not visible to anyone using an Inter-based PC with either Excel v4.0 or v5.0.

Following this analysis it would be helpful to sit down with Ken Bowyer and step through a list of scenarios and review the output. Intermetrics ran a series of limited tests on the model and found that the costs did behave in a linear fashion. However, the reasonableness of the cost as data requirements climb warrant further investigation.

It would be helpful if HAIS could explain if the Sites that are specified are necessary for the model to operate properly. Does the model have a provision that will discriminate and show a cost variance for one site over another? If it does, can HAIS identify the multiplier it and explain the rational surrounding it.

From the limited amount of testing and analysis performed on the Interactive Cost Model it appears that is does function within the parameters outlined. This however, does not substantiate the methods, or assumptions that HAIS used to develop the cost model.

4. *Recommended Distribution*

Originator: Approved:

Layne Miller
EOSDIS IV&V Analyst

Frank Rockwell
EOSDIS IV&V Project Manager